Random Matrix Theory and The City: analyse de la matrice de corrélation des “stock returns”

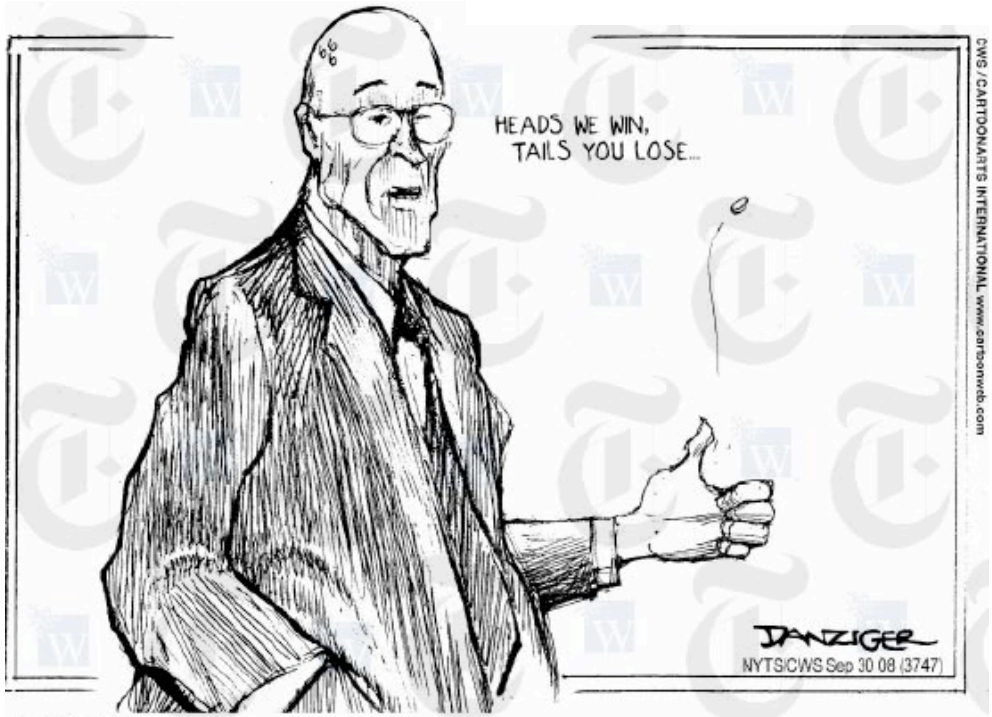
Giulio Biroli

En collaboration avec JP Bouchaud et M Potters, EuroPhys. Lett. (2007), JSTAT (2008), Acta Physica Polonica (2007)

Motivations

Finance et physique? Quelles techniques?
Quelles approches?

Au début la seule chose que je savais est qu'il s'agissait de théorie de probabilité du style...



DANZIGER
CartoonArts International / NYTimes Syndicate

Cadre du travail

- Applications de la théorie des matrices aléatoires.
- Statistique des extrêmes.
- Systèmes désordonnés.

Mini “Crash Course” en finance

- Prix de l’action i : $p_i(t)$
- Le “stock return” est $X_i^t = \frac{p_i(t+\tau) - p_i(t)}{p_i(t)} \simeq \log p_i(t+\tau) - \log p_i(t)$
- $X_i^t(\tau)$ est une marche aléatoire pour des τ d’ordre de plusieurs semaines.

  $$P(X_i^t(\tau)) \propto \frac{1}{X_i^t(\tau)^{1+\mu}} \qquad \mu \simeq 3-5 \qquad \tau \simeq 24h$$

  (hypothèse du marché efficient, Bachelier, Fama,...)
- L'évolution du prix n’est pas tout à fait aléatoire...
- Problème: séparer le signal (vraie évolution dynamique) du bruit.

Matrice de Correlation

$$C_{i,j} = \frac{\langle X_i^t X_j^t \rangle}{\sigma_i \sigma_j}$$

ou σ_i^2 est la variance de X_i^t

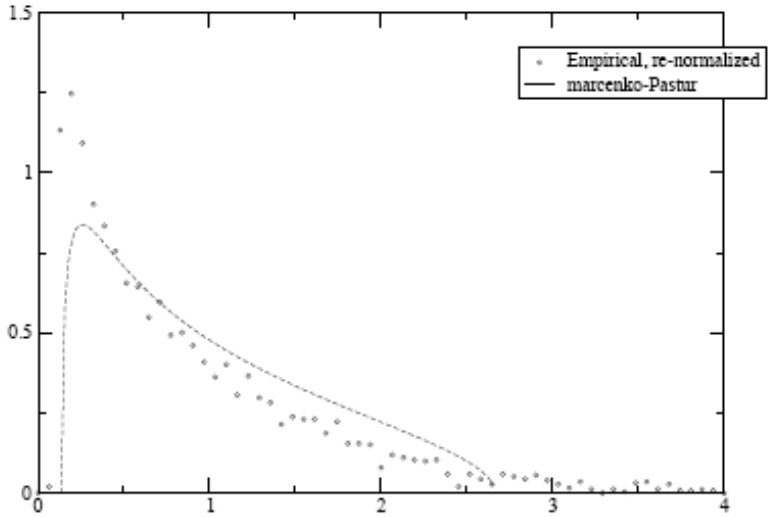
Elle joue un rôle crucial dans l'analyse théorique et aussi dans les stratégies d'investissement.

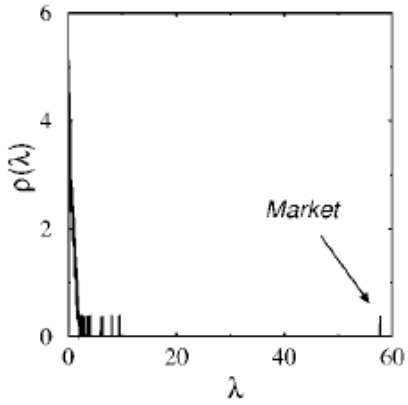
En pratique on mesure

$$E_{ij} = \frac{1}{T} \sum_t \frac{X_i^t X_j^t}{\sigma_i \sigma_j}, \qquad \sigma_i^2 = \frac{1}{T} \sum_t \left(X_i^t\right)^2$$

DOS

N=S&P 500 ;
T=2000-2004





Wishart Ensemble: $E_{ij} = \frac{1}{T}\sum_{k=1}^{T} X_i^k X_j^k$ where $\langle X_i^k X_j^l \rangle = \delta_{ij}\delta_{kl}$

Est-ce que tout ce qui “sort” de la DOS de Marcenko-Pastur est significatif ? En particulier les plus grandes valeurs propres ?

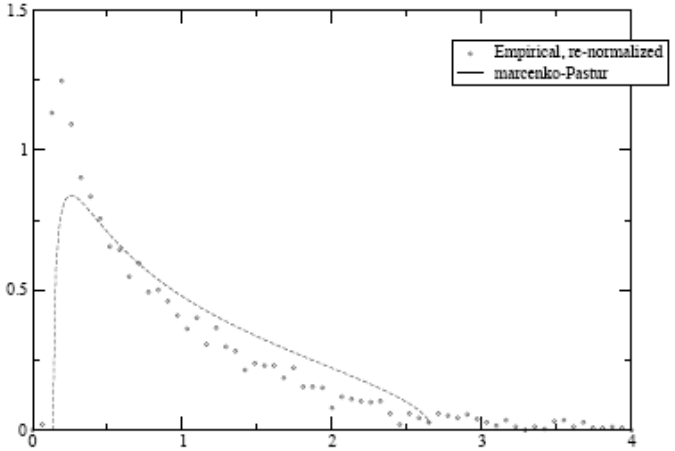
Plus grande valeur propre : mode marché $v_i^M \simeq \frac{1}{\sqrt{N}}$

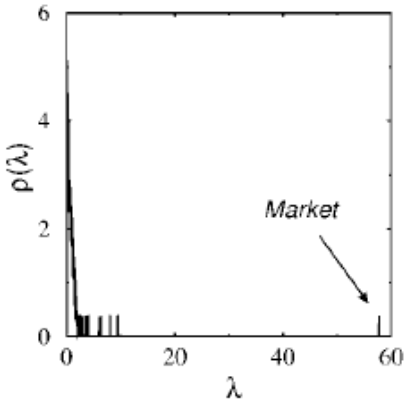
Et les autres modes? Au moins les plus grands?

Pour diminuer le risque dans les investissements le “portfolio” est orthogonal aux modes les plus grands.

DOS

N=S&P 500 ;
T=2000-2004



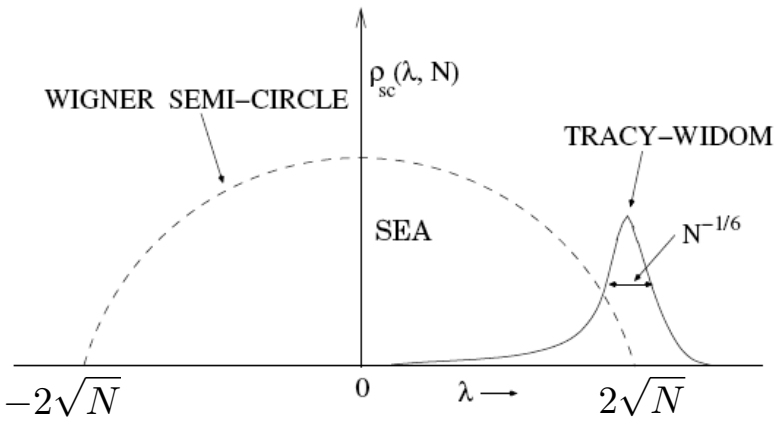


Wishart Ensemble: $E_{ij} = \frac{1}{T}\sum_{k=1}^{T} X_i^k X_j^k$ where $\langle X_i^k X_j^l \rangle = \delta_{ij}\delta_{kl}$

Est-ce que tout ce qui “sort” de la DOS de Marcenko-Pastur est significatif? En particulier les plus grandes valeurs propres?

Statistique des plus grandes valeurs propres

Problème très étudié récemment GOE, GUE, Wishart,... à cause de son universalité.



$$\langle X_{ij} X_{kl} \rangle = \delta_{\langle ij \rangle, \langle kl \rangle}$$

X_{ij} Gaussiennes

Qu'est se passe-t-il pour des variables telles que $P(X_{ij}) \propto \dfrac{1}{X_{ij}^{1+\mu}}$?

Notre résultat pour GOE

- $\mu > 4 : \lambda_{max} = 2\sqrt{N} + \chi N^{-1/6}$ Tracy-Widom distribution.
- $2 < \mu < 4 : \lambda_{max} = Max_{<ij>}|X_{ij}| \propto N^{2/\mu}$ Fréchet distribution
- $\mu = 4 : \lambda_{max} \geq 2\sqrt{N}$ et fluctue sur la même échelle que le demicercle

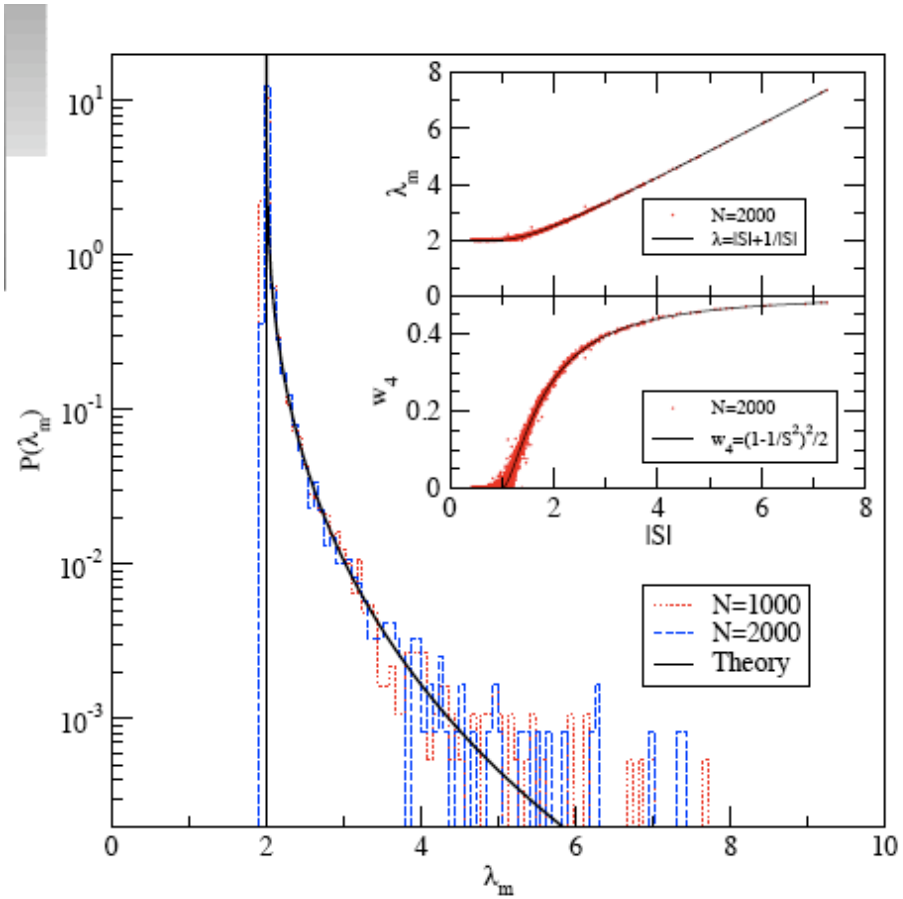
$$s = Max_{\langle ij \rangle}|X_{ij}|$$

$$\lambda_{max} = \left(s + \frac{1}{s}\right)\sqrt{N} \qquad s \geq 1$$

$$\lambda_{max} = 2\sqrt{N} \qquad s < 1$$

Nous avons aussi obtenu la statistique des autres valeurs propres (deuxième, troisième,...) ainsi que les propriétés des vecteurs propres correspondants.

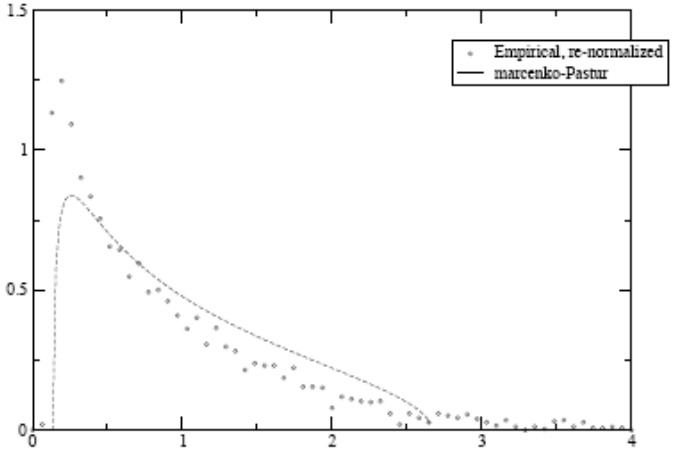
Comparaison avec résultats numériques $\mu = 4$

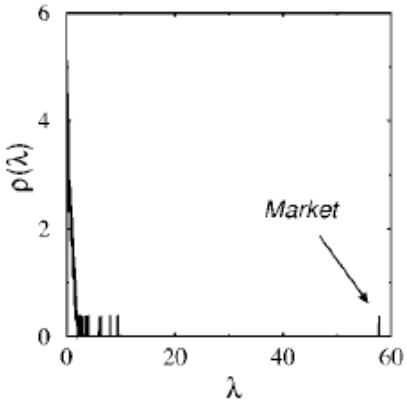


- Nos avons aussi étendus nos résultats au cas Wishart : $\mu = 4$ sépare à nouveau deux classes d'universalité différentes.
- Application aux marchés financiers (mais aussi de façon plus générale) : pas toutes les valeurs propres qui sortent de la DOS de Marcenko-Pastur sont pertinents.
- Certains de nos résultats ont été prouvés rigoureusement par Auffinger, Ben Arous, Péché (2008).
- Universalité des nos distributions limites vs Tracy Widom?

DOS

N=S&P 500 ;
T=2000-2004





Wishart Ensemble: $E_{ij} = \frac{1}{T}\sum_{k=1}^{T} X_i^k X_j^k$ where $\langle X_i^k X_j^l \rangle = \delta_{ij}\delta_{kl}$

Est-ce que tout ce qui “sort” de la DOS de Marcenko-Pastur est significatif ? En particulier les plus grandes valeurs propres ?

Student Wishart Ensemble

- A partir des études empiriques un bon modèle des "stock returns" est: $r_i^t = \sigma_t \eta_i^t$

η_i^t Gaussienne

$\langle \eta_i^t \eta_j^{t'} \rangle = \delta_{ij}\delta_{tt'}$

$$P(\sigma) = \frac{2}{\Gamma(\frac{\mu}{2})} \exp\left[-\frac{\sigma_0^2}{\sigma^2}\right] \frac{\sigma_0^\mu}{\sigma^{1+\mu}}$$

$$P(r_i) = \frac{1}{\sqrt{\pi\mu}} \frac{\Gamma\left(\frac{1+\mu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)} \frac{1}{\left(1+\frac{r_i^2}{\mu}\right)^{\frac{1+\mu}{2}}} \propto \frac{1}{r_i^{1+\mu}} \qquad r_i \gg 1$$

$$\mu \simeq 3 - 5 (3.85)$$

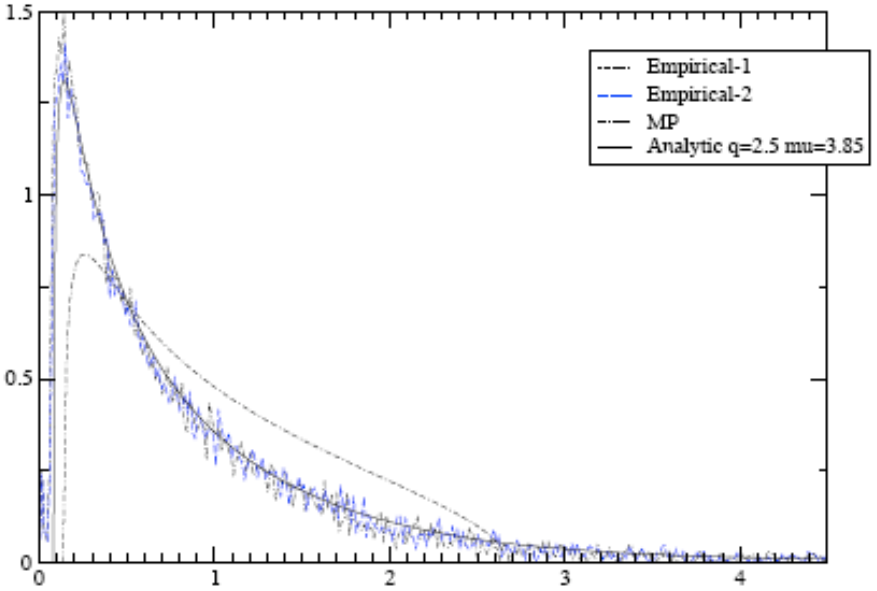
DOS par “free RMT”

- La matrice empirique $W = \frac{1}{T}\sum_t \sigma_t^2 \eta_i^t \eta_j^t$ est une somme de matrices “libres”.

$$B(G(z)) = z \qquad G(z) = Tr(z\mathcal{I} - W)^{-1}$$

- En utilisant que $B(x) - \frac{1}{x}$ est additive pour les matrices libres nous trouvons une équation intégrale sur G(z).

Comparaison avec les données



$\mu = 3.85$

- La distribution décroît comme $1/\lambda^{1+\mu}$, pas de support compact comme Marcenko Pastur.